Assignment 1 Marking Scheme

30 points in total, 10 points per question

Question 1

Part a: need to change binary data to binomial, so df is 437, not 671. 1 mark off if data is not combined to binomial. 1 mark for correct explanation that most n_i's equal to 1 and deviance is not a good measurement of goodness of fit.

NR: The literature seems to be mixed on whether or not this is required. (Even Davison does it both ways in Example 10.18.) If by grouping into covariate classes we end up with many large-ish m_i's, then we have the advantage that the residual deviance can be used as goodness-of-fit. But since this is not the case here, fitting with the original binary data leads to essentially the same results as fitting to the binomial data.

Part c: 1 mark for correct CI, 1 mark for explaining how to calculate this CI statistically, not simply attaching some codes.

The point is that the normal approximation works best on the scale of the linear predictor x_0^{\T}\hat\beta; and this confidence interval is converted to one for $\hat p$ using the inverse logit transform.

Part d and e: 1 mark off if error rates are not calculated and compared.

Part f: All reasonable answers are acceptable as long as the answer is related to the data and error rate calculation above.

Question 2

Most students were able to provide correct calculation and proof.

Part e: 1 mark off if missing the point that Pearson and Deviance test statistic depend on the data only through estimated value of beta and thus fail to evaluate the goodness of fit.

Question 3

Part a: 1 mark off if the reference does not follow any citation style.

Part b and d: a mark off for each part if answered without any explanation.

Part c: 1 mark for study population, 1 mark for population of interest.

Part f: 1 mark off if no statistical method is described for the main analysis.